

DOCUMENT RESUME

ED 300 383

TM 012 190

AUTHOR Neiminen, Gayla; Uguroglu, Margaret
TITLE West Northfield District 31 General Evaluation Model
and Implementation Guide.
INSTITUTION Institute for Educational Research, Glen Ellyn,
IL.
PUB DATE Jun 86
NOTE 26p.
PUB TYPE Guides - Non-Classroom Use (055)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Educational Assessment; Elementary Secondary
Education; Evaluation Methods; *Models; *Program
Evaluation
IDENTIFIERS *West Northfield District 31 IL

ABSTRACT

The evaluation model presented here was developed by the West Northfield (Illinois) District 31 Evaluation Committee in collaboration with consultants from the Institute for Educational Research. Based on a review of the literature prepared by the Institute, the guide elaborates on the steps to be taken in conducting an educational program evaluation. The guide divides program evaluation into the following phases: (1) identification of the purpose of the evaluation; (2) description of the current state of the program; (3) review of pertinent literature and survey of current practices; (4) selection of an appropriate evaluation strategy; (5) collection and analysis of data; and (6) summary of findings and recommendations. This guide is intended to be exhaustive of evaluation procedures and activities; thus, it is up to evaluators to select those activities (phases) that are relevant to their evaluation. Selection of evaluators, role of evaluators, and use of control groups are also covered. (TJH)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED300383

WEST NORTHFIELD DISTRICT 31

GENERAL EVALUATION MODEL

AND

IMPLEMENTATION GUIDE

Presented to:

Dr. Robert Lockett

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- ☒ This document has been reproduced as
received from the person or organization
originating it.
- ☐ Minor changes have been made to improve
reproduction quality.

- Points of view or opinions stated in this docu-
ment do not necessarily represent official
OERI position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

SANDRA K. CUNNINGHAM

June 1986

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

Presented by:

Dr. Gayla Nieminen
Dr. Margaret Uguroglu

© The Institute for Educational Research
793 North Main Street
Glen Ellyn, Illinois 60137

TM 012 190

2

BEST COPY AVAILABLE

WEST NORTHFIELD 31 GENERAL EVALUATION MODEL

Name of Program: _____

Evaluation Committee:

Chair: _____

Position: _____

Members: _____

Position: _____

- I. Determine the purpose of the evaluation.
 - A. What is the overall goal?
 - B. What specific questions must be answered?
- II. Describe the current state of the program.
 - A. What is the program philosophy?
 - B. What are the overall program goals?
 - C. What are the content and behavior objectives?
 - D. What formal activities are carried on?
 - E. What is the pertinent history of the program?
 - F. What is the cost of the program?
- III. Review the literature/determine current practices.
- IV. Select an appropriate evaluation strategy.
 - A. Which groups will be evaluated?
 - B. What type of data will be collected?
 - C. What types of comparisons/analyses are needed?
 - D. What criteria will be used for judgment?
 - E. What is the schedule that will be followed?
- V. Collect and analyze data.
- VI. Summarize the findings and develop recommendations.

WEST NORTHFIELD DISTRICT 31

GENERAL EVALUATION MODEL AND IMPLEMENTATION GUIDE

Overview of this Guide

Many models have been described for different evaluation purposes. The model presented here was developed by the West Northfield District 31 Evaluation Committee in collaboration with consultants from the Institute for Educational Research (IER), and is based on a review of the literature prepared by IER. This implementation guide was developed by Dr. Margaret Uguroglu and Dr. Gayla Nieminen of IER, drawing upon the references listed at the end. These references also provide further explanation for the interested reader.

This guide elaborates on the steps to be taken in conducting a program evaluation, based on this model. While the exact procedures to be followed when accomplishing such an evaluation will vary with the setting, evaluation in general includes the tasks and follows the sequence described here. The guide divides program evaluation into the six phases specified by the model, each corresponding to a major task to be accomplished.

It is important to realize, however, that EVERY step specified here will NOT necessarily be carried out and EVERY question need NOT be answered during a given evaluation. This guide is intended to be exhaustive of evaluation procedures and activities, and it is up to the evaluator(s) to select those activities relevant to the planned evaluation.

- Phase I: Identify the Purpose
- Phase II: Describe the Current State of the Program
- Phase III: Review Literature/Survey Current Practices
- Phase IV: Select an Appropriate Evaluation Strategy
- Phase V: Collect and Analyze Data
- Phase VI: Summarize Findings and Recommendations

Note: It is recommended that the evaluator read the accompanying literature review entitled Evaluation Models in order to get an overview of crucial issues before beginning the evaluation process.

Purpose of Evaluation

Evaluation is the process of determining to what extent objectives have been met. This determination may apply to objectives set for programs, students, teachers, or use of facilities. With the rising pressure of managing scarce resources, determining effectiveness has become crucial. The move toward accountability of effectiveness has made evaluation take on new meaning.

Selection of Evaluators

Selecting evaluators can be a difficult task. Consideration must be given to competence, personal qualities, and the purpose of the evaluation. If an evaluator is familiar with a program or with similar programs, the evaluation can be conducted more efficiently, as time will not be spent educating the evaluator. Technical expertise regarding specific programs as well as in the area of statistics may also be important if quantitative methods are chosen. More effective work results from evaluators who are objective, trusted by administrators, and genuinely interested in improving the program being evaluated.

For West Northfield District 31, using a committee of evaluators instead of only one can be an especially useful approach. A committee can provide for a more complete and balanced evaluation, and can assure a range of needed expertise. In this guide, however, the term "evaluator" is used throughout to signify either one person or a committee conducting the evaluation.

Role of Evaluators

The evaluator of a program has responsibility for producing an accurate description of the program--complete with measures of its effects--that summarizes what has transpired during a particular time period. Results from an evaluation, usually compiled into a written report, can be used for several purposes:

- * To document for a funding agency that the services that were promised by the program's planners have indeed been delivered
- * To assure that a lasting record of the program remains on file
- * To serve as a planning document for people who want to duplicate the program or adapt it to another setting
- * To determine if a program is in need of modification.

PHASE I: IDENTIFY THE PURPOSE

The focusing or general planning phase of a program evaluation is crucial and often difficult. It is a time when the evaluator reads important documents, talks to people, and thinks through the whole evaluation. The job of the evaluator is to collect, digest, and report information about a program to satisfy the needs of one or more audiences. The audiences in turn might use the information for either of three purposes:

1. to learn about the program;
2. to satisfy themselves that the program they were promised did indeed occur and if not, what happened instead;
3. to make decisions about continuing or discontinuing, expanding or limiting the program, generally through giving or withholding funds.

In order to be useful, an evaluation must be focused very clearly on the purposes of its intended audience. If a decision hinges on the findings, the first job is to find out what the decision is. Then the evaluator must ensure that the appropriate information is collected and reported to the correct audiences. The evaluator must decide what information will be needed so that the evaluation is useful to the intended audiences, accurate in its description of the program, and efficiently carried out.

The evaluator will begin the descriptions of the purposes of the evaluation, including the decisions to be made and the audiences, by considering answers to the following questions:

1. Why is this evaluation being conducted? What is the overall goal? What decisions will be based on the evaluation?
2. What specific questions must be answered?
3. Who wants to know about the program? That is, who is the evaluation's audience? Consider the following possible audiences:

Teachers
Administrators
Counselors or department heads
District Personnel
School Board
Parents
Community in general
Superintendent
State Department of Education
Federal Personnel

4. Has one of the evaluation's audiences, such as a Federal or State funding agency, stated specific requirements for this evaluation? For example, is it a requirement to use particular tests, to measure attainment of particular outcomes, or to report on special forms?
5. When must the findings be reported to each of the relevant groups? What is the absolute deadline for the earliest evaluation report?

It is recommended that the answers to the above questions be summarized in writing, and saved for inclusion in the evaluation report.

PHASE II: DESCRIBE THE CURRENT STATE OF THE PROGRAM

This description should be drafted at this point in the evaluation, because it will ensure, from the beginning, a clear grasp of the program including its purposes and critical components. It will also mean less work later, particularly at the busy time when the data have been collected and must be analyzed.

1. Obtain written documents that describe the program, including:
 - Lists of goals and objectives
 - Descriptions made by the staff or others, such as memoranda, meeting minutes, or newspaper articles
 - Curriculum guides for program materials
 - A description of the administrative and staff roles played by those involved in the program
 - The program's budget, particularly the part that mentions the evaluation
 - Descriptions of the program's history
 - District-wide year-end assessment instrument
 - Past evaluations for this or similar programs
 - Program proposal written for a funding agency
 - Results of a needs assessment whose findings the program is intended to address
 - Written state or district guidelines about program processes and goals to which this program must conform
2. In reviewing the documents collected, the evaluator will want to note general impressions of the program, its context, and staff. The evaluator should also be alert for the following details:
 - The philosophy or point of view of the program
 - The program's major general goals

Specifically stated objectives, such as percents of students achieving certain goals, and deadlines for reaching certain objectives

The number and distribution of sites involved

Published curricula used for the program

Curriculum materials constructed in-house

Schedules and lesson plans/descriptions

Staff roles and responsibilities

Student evaluation plans

Teacher evaluation plans

Program evaluation plans

Examples of similar programs that planners intend to emulate

The needs of the community or of the constituency which the program is intended to meet--whether these have been explicitly stated or seem to implicitly underlie the program

3. Following this analysis of documents, the evaluator should check the description of the program against the impressions of the program's planners and staff. In addition, others might be contacted whose actions, opinions, and decisions will influence the success of the evaluation and the extent to which the information collected will be useful and used. This might include:

Evaluators who have worked with this particular program or programs like it. They will have valuable advice to give about the information to collect, how, and from whom.

School or district personnel not directly connected with the project, but whose cooperation will help carry out the evaluation more efficiently or quickly.

Influential parents or community members whose support will help the evaluation go more smoothly

If key people are too busy for this discussion, the evaluator can send a memorandum describing the evaluation, what is needed from the person, and when.

IF POSSIBLE, THE EVALUATOR SHOULD OBSERVE THE PROGRAM IN OPERATION.

4. Prepare a written description of the program which sets the program in context. The amount of detail presented here will depend again upon the audience for whom the evaluation report is prepared. If the audience has no knowledge of the program, it must be fully described. If, on the other hand, the evaluation report is mainly intended for internal use and its readers are likely to be familiar with the program, this description can be fairly brief, setting down information "for the

record." When written, this draft should be circulated to program personnel for comment.

The program description might typically include:

Pertinent History of the program

- a. Where was the program implemented? What sort of communities? How many people--parents, teachers, students--did it affect? What special groups?
- b. How did the program get started?
- c. Was a formal or informal needs assessment conducted, and if so, what were the results?
- d. Were there community, state, or national demands which led to the program? Or were there legal requirements? Opportunities for funding? Teacher initiatives? District requirements?
- e. Were changes made in the philosophy, goals and objectives, activities, or size of the program over time?
- f. When was the most recent evaluation of this program? What were the findings and recommendations?

Goals of the program

- a. What was the program designed to accomplish?
- b. What goals and objectives were set? What were the priorities?

Type of Program

- a. Basic academic (reading, arithmetic, social studies, science)
- b. Special area academic (art, music, health, home economics)
- c. Special education
- d. Non-academic (physical education)
- e. Extracurricular
- f. Student support (counseling)
- g. Other (maintenance, cafeteria, bus service)

Students involved in the program

- a. For what age, grade level, ability level is the program appropriate?
- b. On what bases are students/classrooms/schools to be selected for the program?
- c. Are participants to remain in the program for its duration? If not, what criteria would determine the time of their entrance or exit?

Program characteristics

- a. What materials are used, and how? Are they purchased or produced in-house?
- b. What activities are included in the program?
- c. What resources--funds, physical facilities, transportation--are available, and who provides them?
- d. To what extent does this section describe the program as it was supposed to look as opposed to how it does look?

Program staff

- a. How many teachers, administrators, consultants, secretaries, clerks, aides, volunteers, parents, and other adults are active in the program? What are their roles?
- b. Are they required to have special training or credentials?
- c. Do they receive special training in connection with the program?
- d. How much time do they devote to the program?

Program costs

- a. What were the start-up costs? Fixed costs?
- b. What would the money have been used for had it not been used in the program?
- c. Are there qualitative non-money costs, such as extra fatigue due to the presence of the program?
- d. Are there opportunity costs, that is, good things foregone in order to implement the program?

- e. Are there negative effects of the program on those not in the program, such as resentment, increased class size, etc?

PHASE III: REVIEW LITERATURE/DETERMINE CURRENT PRACTICES

This is an optional activity, which might be especially appropriate when a change in the program is being considered. In this case, the district may be interested in learning what the "best practices" are in the program area, and this information can be obtained from published research or program descriptions, or from other school districts. In addition, reviewing the literature or talking to directors of other similar programs may provide information on the most important variables to be measured during the evaluation.

One other source for similar district programs can be the history of the current program. In reading the pertinent documents and in talking to program staff, the evaluator may find references to model programs which stimulated the development of the program under evaluation, and which could be contacted for further information.

Outside contractors may also provide this information.

PHASE IV: SELECT AN APPROPRIATE EVALUATION STRATEGY

Carrying out an evaluation in any area of education is a matter of collecting evidence to demonstrate the effects of a program or one of its subcomponents. The program's objectives and the purposes of the evaluation will help the evaluator to make broad decisions about what to look at. Four general aspects of a program might be measured as part of the evaluation:

1. Context characteristics

Educational programs take place within a setting or context, the givens for the program. These may include such things as class size, style of leadership in the school, district organization, time frame within which the program must operate, number of classrooms per site, or school budget.

2. Student/participant characteristics

Student characteristics include such things as age, sex, socioeconomic status, language dominance, ability, attendance record, and attitudes.

3. Characteristics of program implementation

Program characteristics are, of course, its principal materials, activities, and administrative arrangements.

4. Program outcomes

The evaluator will almost always want to measure the extent to which goals have been achieved. The evaluator should in general focus the evaluation on announced goals, but should be careful to include the possible wishes of the program's larger constituency--for example, the community--in formulating the yardsticks against which the program will be held accountable.

Beyond these general guidelines, decisions regarding exactly what information to collect will be situation-specific. Every program has distinctive goals and every situation makes available unique kinds of data. Though there is no simple way to decide what specific information to collect, or what variables to look at, here are some general rules to follow:

- a. Focus data collection where program effects are most likely to be noticeable.
- b. Try to collect a variety of information, as data from a variety of sources is stronger and more valid than that from only one source.
- c. Measure what the audience(s) will most likely look for when they receive the report.
- d. Try to develop creative and credible ways to detect achievement of program objectives.

The heart of the evaluation lies in selecting and assessing the outcomes, activities, and costs that might be measured to gauge the success of the program. From all the items possible to measure, it is necessary to narrow down the list to the ones that are most crucial and easily measured within the existing time and money constraints.

The following steps should be undertaken:

1. Decide what must be measured.

In order to do this, it will be necessary to return to the program's objectives, activities, and costs, which were listed in Phase II. All of these could potentially be assessed as part of the program evaluation, but limits in the time and money available will undoubtedly make such a comprehensive evaluation impossible to fulfill.

It is the job of the evaluator to determine which objectives and activities are crucial to the evaluation. This decision will be based on the purpose of the evaluation, and the use to which the documentation will be put. For example, data collected for the purpose of accountability would focus on those activities, administrative changes, or materials that are either specifically required by the program funders or have been put forward by the program's planners as major vehicles for producing its beneficial effects.

The amount of detail with which the characteristics must be described will depend, in turn, on how precisely planners or funders have specified what should take place. The level of detail should equal or exceed the specificity of the program plan, at least when describing the features that the staff consider most crucial to producing program effects.

Three strategies can be useful in achieving the goal of a comprehensive but cost-efficient evaluation. The first involves finding useful information which is going to be collected regardless of the evaluation. For example, the evaluator should find out which tests are given as part of the program or routinely in the setting; should look at the teachers' plans for assessment; and should look at records from the program or at reports, journals, and logs which are kept. It may be that evidence of the achievement of some of the program's objectives can be inferred from these.

A second strategy is to find someone who will collect information. The evaluator might persuade teachers to establish record-keeping systems that will benefit both the evaluation and the teachers' instruction. A local high school or college student might be hired to collect data.

Finally, a cost-effective but useful strategy is to use sampling procedures. They will cut down the time spent administering and interpreting any one measure. Choosing representative sites, events, or participants on which to focus, or randomly sampling groups for testing, will usually produce information as credible to the audience as if the evaluator had looked at the entire population of people served.

Collecting a variety of information gives the advantage of presenting a thorough look at the program. It also gives the evaluator a good chance of finding indicators of significant program effects and of collecting evidence to corroborate other findings.

2. Choose measurement instruments.

The instruments chosen must be appropriate to the variables to be measured and to the overall purposes of the evaluation. Following is a list of some of the instruments that could be used:

- Checklists
- Interviews
- Achievement tests
- Observation schedules
- Questionnaires
- Written records
- Attitude rating scales
- Sociometric instruments

The evaluator will need to select an instrument for each of the variables to be measured which were listed in part A above. This will undoubtedly create a very long list of instruments, which can be trimmed by re-examining the list of variables and determining which can be measured in combination by using the same instrument.

3. Plan the construction or purchase of measurement instruments.

This may be a task which the evaluator could delegate. Like the review of the literature (Phase III), this task could be accomplished by an experienced person. Such a person might be available in the district (school psychologist? department head?), through a local college (education/psychology/sociology faculty or graduate student?) or through an outside consultant. This person can purchase or develop instruments with school personnel, and can also assist in administering the instruments and with the following two steps in this phase. It will be up to the evaluator to build a schedule and set deadlines for these activities.

The following issues should be considered:

What published tests or instruments are available for the purpose and age level needed?

Of those published instruments, which is the most reliable and valid, and easiest to administer?

How best can reliability and validity be assured for the instruments which will be developed in-house or by consultants?

Published instruments should be ordered immediately upon their selection, to ensure that they arrive in time for their use. Those made in-house or by

consultants should be piloted, debugged, and evaluated for technical quality (reliability and validity).

The evaluator must be certain that scoring materials are available for each instrument. If the instrument has a selected-response format, such as multiple-choice, true-false, or Likert scale, scoring key or template will be needed. If it has an open-ended format, the evaluator or the test-constructor must develop a set of criteria for scoring, or a way of categorizing and coding questionnaire or interview responses.

4. Choose an evaluation design for administering instruments.

An evaluation design is a plan of who will receive each measurement instrument which has been selected for administration, and when. Some possible designs are listed below:

True control group

This design is ideal for choosing between the evaluated program and another program, since it shows exactly how the evaluated program (Program A) measures up against an alternative (Program X). In order to implement this design, both programs should be running at the same time.

Procedure

- a. Identify all students or classes who can participate in either program.
- b. Pretest this group.
- c. Randomly divide the students or classes into two groups, and give one group Program A and the other group Program X.
- d. Make sure that:

There is as little difference as possible in what happens to the two groups apart from the Program A/Program X difference--that is, avoid confounding variables.

Program A and Program X stay distinct, not sharing or joining forces--that is, they should avoid contamination.

- e. Posttest both groups simultaneously.

True control group with posttest only

This design is identical to the previous design, except the pretest--step 2--is omitted. This design is useful when the experience of the pretest might itself interfere with program effects. This might be the case, for example, if a questionnaire were administered at the beginning of a program intended to change attitudes, thereby flagging for participants the program's intended effects. A pretest might also be unavailable or inconvenient to give.

Randomization in this design assures equivalence of the groups receiving the two programs.

Non-Equivalent Control Group

This is a good design for choosing between Program A and some other program which will also be in operation during the period of the evaluation. This design makes possible a comparison of two groups which are similar even though they were not randomly assigned and therefore not equivalent as in the previous designs.

Procedure:

- a. Find a similar group of students, at the same grade level who will be given Program X in the same time period that a group will be given Program A.
- b. Pretest both groups of students, and collect information on both groups concerning characteristics which might affect their reactions to Programs A or X.
- c. Investigate differences in what happens to the group getting Program A and the group getting Program X if these differences are likely to affect results.
- d. Posttest both groups.

Single Group Time Series

A series of measurements made during and/or after the program gives a good picture of its impact on the outcomes measured. This design requires scores collected from the same group on several occasions before the program begins and on several occasions during and after the program. The scores should be based on the same measuring instrument.

Time Series with Non-equivalent Control Group

This design involves making measurements exactly as in

the previous design. But here, two groups of students are measured regularly--the group receiving Program A, and the other, a non-equivalent control group, not randomly assigned.

Before-and-After, Design Allowing Informal Comparisons Only

The first two designs described here are experimental designs because both the Program A and a comparable or control program, Program X, are implemented to see which produces better results: a little experiment is run. If only Program A can be measured, then the only way to interpret the results of the measurements will be by comparing pretest and posttest results, or by making informal comparisons based on published data, school records, or predetermined standards. One or more of the following comparisons might be made:

Comparison of pretest and posttest results for Program A

Comparison of Program A results with results from a national sample of students, the norm group for a standardized test.

Comparison of Program A results with results obtained by last year's students at the same grade level.

Comparison of Program A results with prespecified criteria; for example, "80% of the students will achieve 75% of Grade 2 math objectives." Such criteria may have been set up by teachers, the community, or the district; or they may have been specified in a program plan.

The evaluator will want to collect data in the form which can be most easily presented and analyzed. A discussion with the person selected in part 3 of this phase can be helpful in examining the match between the analyses planned and the instruments selected, and in identifying straightforward analyses which might be performed. If the planned data analysis requires mechanical or computer processing, this must also be arranged and scheduled.

5. Choose a sampling strategy for administering instruments.

Sampling plans make data collection more manageable. If the program to be evaluated is a large one with many participants at diverse sites, it will not be possible to measure all the members of each group, or to look at implementation at every site. In this case, it will be necessary to sample people to test, question, or observe, and sites on which to focus may have to be sampled as well.

Choice of which sites to measure depends on the questions which the evaluator is attempting to answer. If the program varies considerably from site to site, representatives of each site will have to be included in the data collection. If the program is intended to be uniform across sites, then sites can be chosen based on other criteria, such as the length of time the program has been in operation at the site, the amount of experience of teachers or other participants, the characteristics of neighborhood involved, or the amount of funds allotted to the program.

In order to produce a representative description of the program, it will be necessary to find out the optimal or most typical times to test students, question people, and observe program implementation at each site. This information should be kept in mind when scheduling the dates of data collection.

6. Estimate the cost of the evaluation.
 - a. Compute a cost-of-staff per unit time figure for each job position occupied by someone who will work on the evaluation.
 - b. Estimate how long people in each position will spend on the evaluation.
 - c. Estimate the evaluation's cost.
 - d. Trim.

After decisions have been made, the evaluator should prepare a description of the planned evaluation and should present it to the program personnel. The evaluator should elicit from the program staff their agreement that the evaluation will provide a fair measure of their program.

PHASE V: COLLECT AND ANALYZE DATA

1. Administer instruments, score them, and record data.

Instruments can be scored as soon as all of the results of one type come in. The evaluator must decide, based on the objectives of the evaluation, whether results will be needed in the form of a score for each examinee, a mean score for each class, or percentage results for each test item. A summary sheet can then be constructed on which to record summarized data for a student, class, or other relevant group. The scoring and summary recording constitute another task which the evaluator may wish to delegate.

2. Do the analysis

Based on the decisions made earlier, the evaluator will need to summarize scores to produce frequencies, percents, measure of central tendency (such as the mean), and indicators of score variability (such as the standard deviation), to fit the evaluation plan. Because the area of data analyses can be relatively complex, it is recommended that the evaluator who is inexperienced in this area refer to a handbook, such as How to Measure Program Implementation or How to Calculate Statistics. These booklets, cited at the end of this Guide, will provide information on graphing, performing tests of statistical significance, and calculating measures of relationship such as correlation.

Whether or not the evaluator can perform statistical tests depends on the design of the evaluation and the measurements made. As a general rule, if an average score per participant or an average response per question was calculated, and if this average for the group can be compared with the average from another group, then a statistical test is appropriate.

Statistical tests are not essential to a good report. However, doing them can make the conclusions stronger.

3. Graph the data.

An optional task, graphing data can allow the evaluator to look for trends and regularities in the results. At times the graph, if it shows a very clear trend, will be all that is needed to support the report conclusions. Results can be graphed even when all the evaluator intends to report is the average scores per group. In this case, the average score for each group could be graphed according to relevant characteristics such as grade level, sex, SES, or ability. Graphing can help uncover much information that the evaluator may otherwise overlook.

When each graph and statistical test is completed, the evaluator should examine it carefully and write a one- or two-sentence description which summarizes the conclusions from reading the graph and noting the results of the analysis. The graphs and summary sentences which give the clearest picture of the program's impact should be saved and used as a basis for the results section of the report.

PHASE VI: SUMMARIZE FINDINGS AND MAKE RECOMMENDATIONS

This part of the guide presents a detailed but standard outline for an evaluation report. Each entry in the report

outline is accompanied by a set of questions intended to clarify the outline section. Answering the questions in sequence will produce a working draft of the evaluation report. The outline presented is rather formal, and is intended to be exhaustive of the type of information that can be conveyed to an evaluation audience. The evaluator may wish to follow the outline as it is, or may use it more informally as a checklist of contents to include.

An Evaluation Report Outline

Front Cover

The front cover should provide the following identifying information:

- Title of the program and its location
- Name of evaluator(s)
- Name(s) of people to whom the evaluation report is to be submitted
- Period covered by the report
- Date the report is submitted

Section I. Summary

The summary is a brief overview (approximately 1-4 pages) of the evaluation report, explaining why the evaluation was conducted and listing its major conclusions and recommendations. It sets the stage for reading the report and also can be used for those who do not need all the details of the full report. Although the summary is placed first, it is the section that is written last.

Typical Content

What was evaluated?

Why was the evaluation conducted?

What are the major findings and recommendations of the evaluation?

And also, if space permits:

Were there decisions to be made on the basis of the evaluation? If so, what decisions?

Which audiences does the report address?

What were the major constraints, if any, under which the evaluation had to be carried out?

Section II. Background Information

This section was prepared as part of Phase II. Refer to pages 6 - 8 for the content of this part of the report.

Section III. Description of the Evaluation Study

Typical Content

Purposes of the evaluation

Evaluation design

Outcome and/or implementation measures, including:
instruments used
data collection procedures

This section was also drafted earlier in the evaluation process. The first part, describing the purposes of the evaluation, was prepared as a part of Phase I, and the parts of this section which cover the evaluation design and the measures were also written earlier, as a part of Phase IV. The reader may refer to pages 4 - 5 for the material to be included regarding purpose, and to pages 10 - 14 for material to be included regarding design and measures.

Section IV. Results

The section presents the results of the various measurements described in Section III. If the measurements were relevant, reliable, and valid, these results constitute "hard data" about the program. In addition, this section might also include some "soft data"--anecdotal evidence or testimonials about the program. Although "soft data" may not be regarded as conclusive because of a lack of representativeness, this kind of evidence enlivens the report and often gets across the program results that cannot be expressed as numbers. While a glowing letter from one parent does not show that the program was a success with all students, it may communicate vividly some of the strengths of the program.

Before this section is written, all data should have been analyzed, recorded in tables, graphed or plotted, and tested for significance where appropriate. Scores from tests are usually presented in graphs and tables showing means and standard deviations for each group. Results of questionnaires are frequently summarized on a copy of the questionnaire itself.

Typical Content

Results of outcome measurements:

How many participants took pretests?

How many of those who took pretests were still in the program group (or control group, if there was one) at the end of the program?

What were the results of pretests? Was there a difference in pretest scores among program and comparison groups?

For each measurement made, what were the posttest results for participants? If a control or comparison group was included, are there differences between the control group and the evaluation group? Are differences statistically significant?

If there was no control group, how much did program group performance change from pretest to posttest?

If there was no control group, to what can results be compared in order to judge their quality? For example, test norms, past performance at program sites, or standards set to reflect different levels of competence may be used for such purposes of comparison.

What were the results of other statistical tests?

Results of implementation measurements

Did the staff deliver the program they had promised?

Was the program implemented as planned and as the audience expected? If not, what happened? Were some components dropped or modified? Were all materials available, and were they used? Was the program given to the students for whom it was planned? Did crucial activities in fact occur?

What did the program finally look like? Will a table describe this well?

If changes occurred in the program, what effects did they have on, for example, staff and student attitudes or efficiency of operation?

How similar was the program from site to site: what variations were allowed?

What did a typical lesson or program exercise look like?

Informal results

Are there available informal comments or narrative summaries of reports which illustrate the findings indicated by the hard data? These might be anecdotal accounts from teachers, parent letters, open-ended questionnaire responses, student essays, or transcripts of informal observations by principals.

Were these testimonials about the program solicited or voluntary?

Section V. Discussion of results

Interpretation of each result could occur in Section IV where it is presented. However, if the program and/or the evaluation is complicated, a separate section for interpreting and discussing the results makes the report clearer. The results should be discussed with particular reference to the purposes of the evaluation listed in Section III.

Typical Content

There are two major issues to be addressed in this section:

1. With what certainty can we say that the program caused the results?

Are there alternative explanations of program results? For example, were the gains made by students perhaps simply the result of normal maturation? The evaluator should attempt here to anticipate and answer arguments against the attribution of program results that might be raised by a skeptic.

2. How good were the results of the program?

How did the program results compare with what might have been expected had there been no program?

If a control group was used, were the program results better than control group results? If the difference was statistically significant, was the size or quality of the difference enough to make it educationally significant? If the difference was not statistically significant, did program results appear promising or in the proper direction?

If there was no control group, how do results compare with whatever standard for judging was established?

Does the staff feel the program would achieve more significant gains were it to be modified or run for a longer period of time? Is there evidence in the data to support this point of view?

What conclusions or recommendations can be made from correlational studies? Do certain program features, for example, seem most effective with certain groups? Are attitudes or achievement related to student, teacher, or parent characteristics?

Section VII. Conclusions and Recommendations

It is best to present this section in the form of a list rather than as a narrative. The recommendations can be the most influential part of the evaluation report. Therefore, the evaluator must be sure to emphasize what is important, and to make clear which conclusions must be tentatively rather than firmly drawn. The evaluator must also take care that this section attends to all the concerns that were described in Section III regarding the purpose of the evaluation.

Typical Content

Conclusions

What are the major conclusions to be drawn about the effectiveness of the program as a whole? Of its various subcomponents? How firm are these conclusions?

Must judgment be withheld regarding some aspects of the program?

Did the evaluation overlook features or effects of the program that should have been considered in order to provide a complete picture of its impact?

Recommendations regarding the program

On the basis of the evaluation, what recommendations can be made concerning the program?

What predictions, hypotheses, or hunches are suggested by the results? What recommendations about further programs or research studies can be made?

Recommendations concerning subsequent evaluations of the program

What instruments served their purpose well and would therefore be recommended for use in subsequent evaluations of this or similar programs?

Should some instruments or evaluation procedures be modified or discarded for subsequent evaluations?

Should components of the overall program be separately evaluated?

If this evaluation has not been able to provide firm answers to some questions raised about the program, what evaluation design could yield such answers?

REFERENCES

- Fitz-Gibbon, C.T. & Morris, L.L. (1978). How to Design a Program Evaluation. Beverly Hills, CA: Sage Publications.
- House, E.R. (1980). Evaluating with Validity. Beverly Hills, CA: Sage Publications.
- McGreal, T. (1983). Successful Teacher Evaluation. Alexandria, VA: ASCD.
- Millman, J. (Ed.) (1981). Handbook of Teacher Evaluation. Beverly Hills, CA: Sage Publications.
- Morris, L.L. & Fitz-Gibbon, C.T. (1978). How to Measure Achievement. Beverly Hills, CA: Sage Publications.
- Morris, L.L. & Fitz-Gibbon, C.T. (1978). How to Measure Program Implementation. Beverly Hills, CA: Sage Publications.
- Morris, L.L. & Fitz-Gibbon, C.T. (1978). How to Measure Attitudes. Beverly Hills, CA: Sage Publications.
- Morris, L.L. & Fitz-Gibbon, C.T. (1978). How to Calculate Statistics. Beverly Hills, CA: Sage Publications.
- Posavac, E.J. & Carey, R.G. (1980). Program Evaluation: Methods and Case Studies. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Weiss, C.H. (1972). Evaluation Research: Methods of Assessing Program Effectiveness. Englewood Cliffs, NJ: Prentice-Hall, Inc.